 Wildlife Research, 1998, **25**, 643–648

Variance estimate for the Allen activity index

Richard M. Engeman[A], *Lee Allen*[B] *and Gary O. Zerbe*[C]

[A]USDA/APHIS/ADC National Wildlife Research Center,
1716 Heath Parkway, Fort Collins, CO 80524-2719, USA.
[B]Department of Natural Resources, Alan Fletcher Research Station,
PO Box 36, 27 Magazine Street, Sherwood, Qld 4075, Australia.
[C]University of Colorado Health Sciences Center, Department of Preventive Medicine and Biometrics, Campus Box B-119, 4200 E 9th Ave., Denver, CO 80262, USA.

Abstract

The Allen activity index, originally developed for monitoring dingo populations, is statistically described as a mixed linear model, from which a variance formula for the index is derived. The resulting formula requires input of variance component estimates, the estimation of which is accomplished using restricted maximum-likelihood estimation. An example is used to demonstrate the calculation of the variance components and their use in the variance formula. Application of the variance formula substantially enhances the quantitative practicality of this useful index of wildlife populations.

Introduction

A frequent problem in wildlife biology occurs when the population and/or density of the animal of interest is impossible to accurately assess with current methods, or the logistical costs of doing such an assessment are prohibitive. The development of an index that tracks changes in the target population within appropriate time and geographic constraints can provide the necessary information to make management decisions. Logistically more complex procedures requiring difficult-to-meet analytical assumptions, such as capture–recapture models, can be avoided. The use of tracking tiles for rats (e.g. Fiedler 1994), scent-post surveys for coyotes (Roughton and Sweeny 1982), and the open-hole method for pocket gophers (Richens 1967) are examples of population-indexing methods. Variance calculations for such indices are usually a result of only the sampling plan and not inherent to index methodology itself.

Allen *et al.* (1996) recently introduced an activity index (AI) for assessing dingo (*Canis lupis dingo*) populations, which is based on observing the number of animal intrusions (number of sets of tracks) on each of a series of tracking plots in the area being assessed. Data are collected from each tracking plot over consecutive days and the mean number of sets of tracks on the plots is calculated daily for the measurable plots (i.e. plots not erased by the elements, cattle, vehicles, etc.). The AI is then formed by calculating an overall mean from the daily means.

Allen and Engeman (1995) demonstrated the versatility and utility of this indexing method for simultaneously monitoring the activity of a variety of species in addition to dingoes as the target species. Its value for monitoring dingoes could lead to application on canid species world-wide. Canids such as coyotes, foxes, dingoes, wolves, jackals and wild dogs are often in conflict with human interests throughout the world, particularly with respect to depredation on livestock, but also for transmission of diseases such as rabies and predation on endangered species (waterfowl, kit fox, ferrets). In particular, interest in indexing coyote populations has existed for some time, including ten consecutive years where the scent-post method (Roughton and Sweeney 1982) was applied in a west-wide coyote survey in the United States. This and other methods have met with varying degrees of utility and success (Knowlton 1984). Its versatility for monitoring multiple species at the same time also holds potential for extensive application. Here, too, the coyote, like the dingo, is a subject of great interest, both for evaluating predator–prey relationships or examining interactions with other canids such as foxes or wolves (e.g. Gese *et al.* 1996).

1035-3712/98/060643

One drawback to the utility of the AI (and many other indices as well) has been the lack of a variance formula. Approximate calculation methods, such as calculating a variance estimate by subsetting the data, provide rudimentary, but potentially biased estimates of variance for the index over all data. Carrying methods such as this forward to resampling plans such as bootstrap or jackknife methods (e.g. Efron 1982) are computationally arduous, but may produce credible variance estimates. Those methods also would be unnecessary and it would be easier for the investigator if a variance formula for AI existed. Here we derive a formula for a variance estimate of AI.

Methods

Data structure

We first simplify the formulation of the AI. As originally defined (Allen *et al.* 1996), data collection was repeated on each plot on consecutive days until the cumulative mean across plots and days changed by less than 10% from the previous day's calculation. In nearly all of our dingo trials over the last five years we have found that this criterion was met after monitoring the tracking plots for four days. Therefore, we presume that the number of consecutive days for monitoring tracking plots will have been fixed in advance as part of the study protocol. We do not specify this optimal period, as different species may be best monitored for different periods, or logistics for some situations may determine the monitoring time frame.

We now formally define in statistical terms the data structure from which the AI is calculated. Assume that p plots will be observed for tracks on each of d days. Let x_{ij} represent the number of sets of tracks found on the ith plot on the jth day. We now write a mixed linear model (e.g. McLean *et al.* 1991; Wolfinger *et al.* 1991) to describe the x_{ij}:

$$x_{ij} = \mu + P_i + D_j + e_{ij}.$$

The term μ is the overall mean number of sets of tracks per plot per day for the area being assessed. D_j is a random effect due to the day on which an observation was made with j = 1, 2, 3 ... d, and d is the number of days the plots are monitored. P_i is a random effect due to the ith plot with i = 1, 2, 3 ... $p_j \leq p$ representing the number of plots contributing data on the jth day. In practice it might be unreasonable to presume that no stations would be rendered unobservable by the elements, or other factors out of the control of the investigator, for each of the d days. Thus, we have allowed the number of plots used in the calculations to differ between days. The e_{ij} represents random error associated with each plot each day.

We need to also make biologically realistic assumptions concerning the distribution of the random effects prior to calculating the variance of AI. Many animals, including the canids for which the AI was originally targeted, roam distances greater than those by which the tracking plots are likely to be separated. Also, plots that are closer together probably share more characteristics that relate to an animal leaving tracks than do more distantly separated plots. Therefore, we do not consider the number of sets of tracks observed on the plots to be independent. Similarly, we cannot consider environmental and climatic conditions to be unrelated across days. Hence, we also do not consider the number of sets of tracks observed on each day to be independent. The e_{ij}, as random observational noise, are considered independent and identically distributed with mean = 0 and variance = σ_e^2.

Variance estimation

The calculation of the AI can now be written in terms of the x_{ij} as

$$AI = \frac{1}{d}\sum_{j=1}^{d} \frac{1}{p_j} \sum_{i=1}^{p_j} x_{ij}$$

Then the variance of the AI is

$$var(AI) = var(\frac{1}{d}\sum_{j=1}^{d} \frac{1}{p_j} \sum_{i=1}^{p_j} x_{ij}),$$

which can be equivalently written as

$$var(AI) = \frac{1}{d^2} cov\left(\sum_{j=1}^{d} \frac{1}{p_j} \sum_{i=1}^{p_j} x_{ij},\ \sum_{j'=1}^{d} \frac{1}{p_{j'}} \sum_{i'=1}^{p_{j'}} \right)$$

$$= \frac{1}{d^2} \sum_{j=1}^{d} \sum_{j'=1}^{d} \frac{1}{p_j} \frac{1}{p_{j'}} \sum_{i=1}^{p_j} \sum_{i'=1}^{p_{j'}} cov(x_{ij}, x_{i'j'})$$

If we let the var(P_i) = σ_p^2 and var(D_j) = σ_d^2, then using the definitions and assumptions given in the subsection on data structure, the covariance structure below follows, with the nonzero elements resulting from the lack of independence among observations:

$$cov(x_{ij}, x_{i'j'}) = \begin{cases} \sigma_p^2 + s_d^2 + s_e^2, & \text{if } i=i' \text{ and } j=j' \\ \sigma_p^2, & \text{if } i=i' \text{ and } j\neq j' \\ \sigma_d^2, & \text{if } i\neq i' \text{ and } j=j' \\ 0, & \text{if } i\neq i' \text{ and } j\neq j'. \end{cases}$$

Substitution into the quadruple summation of the variance formula produces the following result:

$$var(AI) = \frac{\sigma_p^2}{d} \sum_{j=1}^{d} \frac{1}{p_j} + \frac{\sigma_d^2}{d} + \frac{\sigma_e^2}{d^2} \sum_{j=1}^{d} \frac{1}{p_j}$$

Note that if all p of the tracking plots provide observations each day, then this formula simplifies to

$$var(AI)_{equal\ sample\ size} = \frac{\sigma_p^2}{p} + \frac{\sigma_d^2}{d} + \frac{\sigma_e^2}{pd}$$

Estimation of var(AI) requires variance component estimates for σ_p^2, σ_d^2, σ_e^2, which can be produced by applying the x_{ij} as observations in a linear mixed-model structure and using a program such as SAS PROC MIXED or PROC VARCOMP (SAS Institute 1992, 1996, 1997), each with a restricted maximum-likelihood estimation procedure (REML), to produce the variance component estimates.

Example

The data in Table 1 were collected for assessing a dingo population near Mt Owen in south-west Queensland. A 50-km transect consisting of p = 50 plots spaced every kilometre was monitored for d = 4 consecutive days. The average number of sets of tracks per plot were 0.94, 0.82, 1.30, 0.82 for days 1, 2, 3, 4 respectively. The AI index value was calculated as 0.97. Application of PROC MIXED or VARCOMP in SAS produces variance component estimates of σ_p^2 = 0.0124, σ_d^2 = 0.0018, and σ_e^2 = 0.0338. We can use the equal-sample-size formula because all plots were measurable on each of the four days, i.e. $p_1 = p_2 = p_3 = p_4 = 50$ for Days 1–4. Insertion of the above information into the equal-sample-size equation for var(AI) yields:

var (AI) = 0.0124/50 + 0.0018/4 + 0.0338/200 = 0.000867
standard error (s.e.) = 0.029
coefficient of variation (c.v.) = 0.030.

Appendix I presents the code for using SAS PROC VARCOMP and PROC MIXED to calculate the components of variance for this example.

Although the trial design from which our example data came was largely determined by the geography of the station and logisitics, we consider the effects on the results had we only been able to observe plots for three days instead of four, or only observe 38 plots instead of 50 (a similar reduction in observations). The AI for the 25% reduction in days or plots are, respectively 1.02 and 1.07, and the associated variance estimates are, respectively 0.00199 and 0.003616.

Table 1. Number of sets of dingo tracks observed on 4 consecutive days from 50 plots spaced 1 km apart along a transect in south-west Queensland

Plot #	Day 1	Day 2	Day 3	Day 4	Plot #	Day 1	Day 2	Day 3	Day 4
1	1	1	2	6	27	1	0	0	0
2	1	1	3	5	28	0	3	2	1
3	3	1	3	4	29	4	2	0	0
4	1	1	4	1	30	3	0	0	3
5	3	0	4	3	31	5	0	0	0
6	4	0	4	0	32	5	0	2	0
7	1	0	4	0	33	2	1	3	0
8	0	1	0	0	34	4	0	2	1
9	0	0	1	0	35	3	0	0	1
10	0	0	0	0	36	0	2	2	0
11	0	1	1	0	37	0	1	3	1
12	0	2	0	0	38	0	0	0	1
13	0	2	1	0	39	0	0	1	0
14	0	0	2	0	40	0	1	2	1
15	0	1	2	0	41	0	1	1	1
16	0	2	0	1	42	1	1	1	1
17	0	0	0	1	43	1	1	0	4
18	0	2	3	0	44	0	1	0	0
19	0	1	3	0	45	0	0	0	1
20	2	0	2	0	46	0	0	0	1
21	0	4	1	0	47	0	2	1	2
22	1	0	1	0	48	0	1	0	0
23	0	0	0	0	49	0	2	0	0
24	0	0	1	0	50	0	0	1	1
25	0	2	2	0					
26	1	0	0	0	Mean	0.94	0.82	1.30	0.82

These estimates are within 10% of the AI calculated from the full data set, but the variance estimates were 2.3 and 4.2 times larger. As would be expected from the variance component estimates, plots had a greater effect on estimation than days, although both were important.

Discussion

There are several important points to make relative to the derivation, calculation and application of the variance formula for the AI. First, the AI is unusual among activity indices in that its implementation defines a data structure that is well-described by a linear mixed model. Use of the model structure and minimal assumptions concerning the relationship among plots through space and time permitted the derivation of a variance formula that could provide a measure of precision each time an index is calculated.

Beyond the derivation of the variance formula, we also demonstrated current methods (REML estimation) and software (SAS PROC MIXED and VARCOMP) for estimating the variance components that are needed in the AI variance formula from mixed linear models. Many '(old) standard' statistical texts (e.g. Snedecor and Cochran 1989; Sokal and Rohlf 1995) present variance component estimation in the context of method-of-moment estimation from analysis of variance tables. This approach has severe weaknesses (e.g. Searle *et al.* 1992), including the potential for negative variance component estimates. With current capabilities of personal computers, the more appropriate methods for estimating variance components from mixed linear models can be accomplished on the desktop using iterative procedures such as maximum likelihood or the more preferred REML estimation (Searle *et al.* 1992). The text by Searle *et al.* (1992) is generally accepted as the current 'standard' for variance component estimation, while the mixed model discussions in the SAS/STAT manuals (e.g. SAS Institute 1996) provide additional useful reference material.

The variance formula allows the quality of a calculated AI to be assessed on the basis of precision using variance, standard error and coefficient of variation. The calculation of the variance components used in the variance formula also provides the investigator with useful information for planning future studies, as the relative contributions of plot-to-plot variation and day-to-day variation can be examined to optimise the combination of days and plots for the next assessment. Conceptually, this approach is similar to that of Link *et al.* (1994), in the much different context of bird counts, as they examine the effects of variability due to the inexactness of surveying wildlife populations. They conclude that replication of counts within each survey site generally should receive less emphasis than acquiring additional survey sites, although they indicate, as we have, that costs and logistics could be the determining factors in setting up a design. Rather than two, we estimated three sources of variation for the AI variance formula and, for our particular example, we found plots had a greater effect on estimation than days. Although both sources of variation impacted estimation, if a logistical choice had to be made between adding more plots or adding more days of observation, then the addition of more plots would receive greater emphasis.

If the AI is being used to monitor populations within an area at different times, or among different areas, then statistical comparisons of the AI would be of interest, especially when looking at topics such as dingo populations before and after a control program, or populations in areas with and without control. This is easily accomplished by calculating the AIs to be compared and their respective variance estimates, followed by the application of the standard z-test for comparing means, or equivalently, the Wald statistic (e.g. Mantel 1987). The z statistic also can be used to calculate confidence intervals for the AI.

Caughley *et al.* (1977) demonstrated the difficulties and assumptions one must make to produce a variance estimate when the sampling methodology does not provide the theoretical basis from which a variance can be derived directly. Their challenge was even greater due to their use of aerial survey methods that provide an excellent population index to directly estimate the harvested proportion of kangaroo populations. Fortunately, the AI data structure permits a straight-forward variance estimation procedure, although making the jump from an index to actual population estimates also would require additional measurements to develop 'correction' factors. Caughley *et al.* (1977), after conducting a meticulous study to address their objectives, used the information generated to estimate the sampling intensity required to achieve a desired precision for a future estimate. If the day and plot variance components have been estimated from an earlier application of the AI and estimation of its variance, then the number of days and plots required to produce a desired precision for a similar future situation can be estimated by examining the variance formula (equal-sample-size version) as a response surface question with days and plots as the independent variables and the variance as the dependent variable. Obviously, with two independent variables and one dependent variable no single solution would exist, but might be optimised within the constraints of the experimental resources. As indicated previously, it has been our experience that experimental logistics and resources are usually the most important influences on sampling design.

Definition of the observational structure of the AI in terms of a linear mixed model has led to increased quantitative practicality for applying this highly useful (Allen and Engeman 1995; Allen *et al.* 1996) indexing procedure. A formula for the variance of the AI was derived. Next, the variance components needed for the AI variance formula can be estimated accurately through REML procedures applied to the mixed-model data. In addition to their use in the AI variance formula, those variance components can be used to optimise future applications of the AI. Lastly, statistical comparisons among AI obtained at different times or from different places can be conducted.

References

Allen, L., and Engeman, R. M. (1995). Assessing the impact of dingo predation on wildlife populations. In 'Tenth Australian Vertebrate Pest Control Conference'. (Ed. M. Statham.) pp. 72–79.

Allen, L., Engeman, R. M., and Krupa, H.W. (1996). Evaluation of three relative abundance indices for assessing dingo populations. *Wildlife Research* **23**, 197–206.

Caughley, G., Sinclair, R. G., and Wilson, G. R. (1977). Numbers, distribution and harvesting rate of kangaroos on the inland plains of New South Wales. *Australian Wildlife Research* **4**, 99–108.

Efron, B. (1982). 'The Jackknife, the Bootstrap and Other Resampling Plans.' (Society for Industrial and Applied Mathematics: Philadelphia.)

Fiedler, L. A. (1994). 'Rodent Pest Management in Eastern Africa.' (Food and Agriculture Organization of the United Nations: Rome.)

Gese, E. M., Stotts, T. E., and Grothe, S. (1996). Interactions between coyotes and red foxes in Yellowstone National Park, Wyoming. *Journal of Mammology* **77**, 377–382.

Knowlton, F. (1984). Feasibility of Assessing Coyote Abundance on Small Areas. US Department of Interior Fish and Wildlife Service, Denver Wildlife Research Center, Work Unit 909:01 Report.

Link, W. A., Barker, R. J., sauer, J. R., and Droege, S. (1994). within-site variability in surveys of wildlife populations. *Ecology* **75**, 1097–1108.

Mantel, N. (1987). Understanding Wald's test for exponential families. *The American Statistician* **41**, 147–148.

McLean, R. A., Sanders, W. L., and Stroup, W. W. (1991). A unified approach to mixed linear models. *The American Statistician* **45**, 54–64.

Richens, V. B. (1967). The status and use of gophacide. *Proceedings Vertebrate Pest Conference* **3**, 118–125.

Roughton, R. D., and Sweeney, M. W. (1982). Refinements in scent-station methodology for assessing carnivore populations. *Journal of Wildlife Management* **46**, 217–229.

SAS Institute. (1992). 'SAS Technical Report P-229. SAS/STAT Software: Changes and Enhancements.' (SAS Institute: Carey, North Carolina.)

SAS Institute. (1996). 'SAS/STAT User's Guide.' (SAS Institute: Carey, North Carolina.)

SAS Institute. (1997). 'SAS/STAT – Software Changes and Enhancements through Release 6.12.' (SAS Institute: Carey, North Carolina.)

Searle, S. R., Casella, G., and McCulloch, C. E. (1992). 'Variance Components.' (Wiley, New York.)

Snedecor, G. W., and Cochran, W. G. (1989). 'Statistical Methods.' (Iowa State University Press: Ames, Iowa.)

Sokal, R. R., and Rohlf, F. J. (1995). 'Biometry.' (Freeman: New York.)

Wolfinger, R. D., Tobias, R. D., and Sall, J. (1991). Mixed models: a future direction. *16th SAS Users Group Conference* **16**, 1380–1388.

Manuscript received 5 May 1997; accepted 24 April 1998

Appendix 1. SAS code for calculating variance components first using PROC MIXED, and then using PROC VARCOMP, where the data are contained in a file named TRACKS.DAT formatted into 3 columns for plot number, day number, and observed number of sets of tracks

Code for PROC MIXED:

```
data a;
infile tracks.dat;
input plot day tracks;
proc mixed method = reml;
  class plot day;
  model tracks = ;
  random plot day;
run;
```

Code for PROC VARCOMP:

```
data a;
infile tracks.dat;
input plot day tracks;
proc varcomp method = reml;
  class plot day;
  model tracks = plot day/fixed = 0;
run;
```